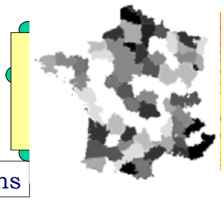
Spatial processes and statistical modelling

Peter Green
University of Bristol, UK
BCCS GM&CSS 2008/09 Lecture 8

Spatial indexing



- Continuous space
- Discrete space
  - lattice
  - irregular - general graphs
  - areally aggregated
- Point processes
  - other object processes

Space vs. time

- apparently slight difference
- profound implications for mathematical formulation and computational tractability

---

Requirements of particular application domains

- agriculture (design)
- ecology (sparse point pattern, poor data?)
- environmetrics (space/time)
- climatology (huge physical models)
- epidemiology (multiple indexing)
- image analysis (huge size)

Key themes

- conditional independence
  - graphical/hierarchical modelling
- aggregation
  - analysing dependence between differently indexed data
  - opportunities and obstacles
- literal credibility of models
- Bayes/non-Bayes distinction blurred

Why build spatial dependence into a model?

- No more reason to suppose independence in spatially-indexed data than in a time-series
- However, substantive basis for form of spatial dependent sometimes slight - very often space is a surrogate for missing covariates that are correlated with location

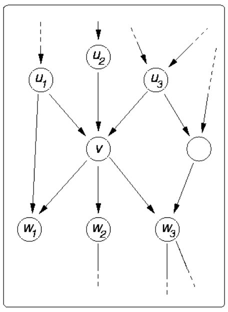
Discretely indexed data

Modelling spatial dependence in discretely-indexed fields

- Direct
- Indirect
  - Hidden Markov models
  - Hierarchical models

Hierarchical models, using DAGs

Variables at several levels - allows modelling of complex systems, borrowing strength, etc.



Modelling with undirected graphs

Directed acyclic graphs are a natural representation of the way we usually *specify* a statistical model - directionally:

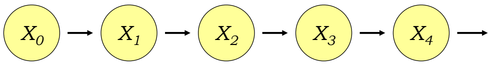
- disease → symptom
- past → future
- parameters → data ……

whether or not causality is understood.

But sometimes (e.g. spatial models) there is *no natural direction*

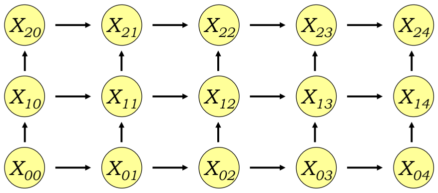
Conditional independence

In *model specification*, spatial context often rules out directional dependence (that would have been acceptable in time series context)



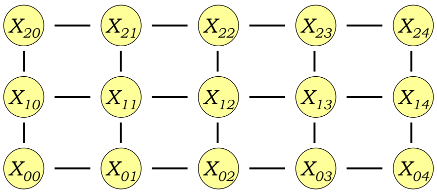
Conditional independence

In *model specification*, spatial context often rules out directional dependence

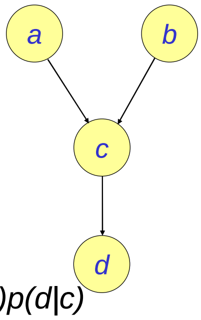


Conditional independence

In *model specification*, spatial context often rules out directional dependence



Directed acyclic graph



in general:

$$p(x)=\prod_{v\in V} p(x_v \mid x_{\mathrm{pa}(v)})$$

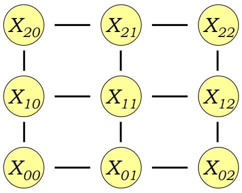
for example:

$$p(a,b,c,d)=p(a)p(b)p(c|a,b)p(d|c)$$

In the RHS, *any* distributions are legal, and uniquely define joint distribution

---

Undirected (CI) graph



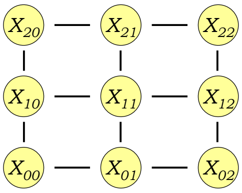
Regular lattice, irregular graph, areal data...

Absence of edge denotes conditional independence given all other variables

But now there are non-trivial constraints on conditional distributions

---

Undirected (CI) graph



Suppose we assume (♣)

$$p(X)\propto \exp\left\{\sum_C V_C(X_C))\right\}$$

clique

then

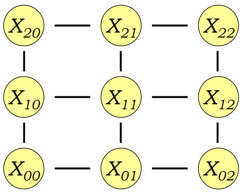
$$p(X_i \mid X_{-i})\propto \exp\left\{\sum_{C\ni i} V_C(X_C))\right\}$$

and so

$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$

The Hammersley-Clifford theorem says essentially that the converse is also true - the only sure way to get a valid joint distribution is to use (♣)

Hammersley-Clifford



A positive distribution $p(X)$ is a Markov random field

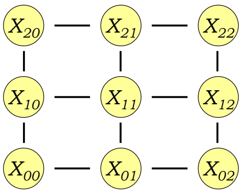
$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$

if and only if it is a Gibbs distribution

$$p(X)\propto \exp\left\{\sum_C V_C(X_C)\right\}$$

- Sum over cliques C (complete subgraphs)

Partition function



Almost always, the constant of proportionality in

$$p(X)\propto \exp\left\{\sum_C V_C(X_C)\right\}$$

is not available in tractable form: an obstacle to likelihood or Bayesian inference about parameters in the potential functions $\{V_C(X_C))\}$

Physicists call $Z=\int_X \exp\left\{\sum_C V_C(X_C)\right\}$ the partition function

Markov properties for undirected graphs

- The situation is a bit more complicated than it is for DAGs. There are 4 kinds of Markovness:
- P – pairwise
  - Non-adjacent pairs of variables are conditionally independent given the rest
- L – local
  - Conditional only on adjacent variables (neighbours), each variable is independent of all others

- G – global
  - Any two subsets of variables separated by a third are conditionally independent given the values of the third subset.
- F – factorisation
  - the joint distribution factorises as a product of functions of cliques
- In general these are different, but F⇒G⇒L⇒P always. For a positive distribution, they are all the same.

Gaussian Markov random fields: spatial autoregression

If $V_C(X_C)$ is $-\beta_{ij}(x_i-x_j)^2/2$ for $C=\{i,j\}$ and 0 otherwise, then

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

is a multivariate Gaussian distribution, and

$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$

is the univariate Gaussian distribution

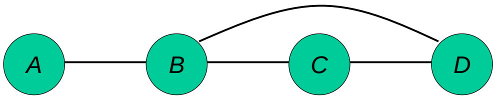
$$X_i \mid X_{-i} \sim N(\sigma_i^2 \sum_{j\in\partial i} \beta_{ij} X_j, \sigma_i^2) \quad \text{where} \quad \sigma_i^2 = 1/\sum_{j\in\partial i} \beta_{ij}$$

$$\begin{array}{c|cccc} & A & B & C & D \\ \hline A & 2 & 1 & 0 & 0 \\ B & 1 & 2 & -1 & 1 \\ C & 0 & -1 & 4 & 2 \\ D & 0 & 1 & 2 & 3 \end{array}$$

Gaussian random fields

non-zero ⇔ $\mathrm{cov}(B,D \mid A,C)$ non-zero

Inverse of (co)variance matrix: dependent case



Non-Gaussian Markov random fields

Pairwise interaction random fields with less smooth realisations obtained by replacing squared differences by a term with smaller tails, e.g.

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

$$= \exp\left\{-\beta\delta(1+\delta)\sum_{i\sim j} \log\cosh(\frac{x_i - x_j}{\delta})\right\}$$

Discrete-valued Markov random fields

Besag (1974) introduced various cases of

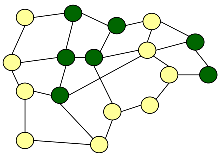
$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

for discrete variables, e.g. auto-logistic (binary variables), auto-Poisson (local conditionals are Poisson), auto-binomial, etc.

Auto-logistic model (X_i = 0 or 1)

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

$$\propto \exp\left\{\sum_i \alpha_i x_i + \sum_{i\sim j} \theta_{ij} x_i x_j\right\}$$



$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$ is Bernoulli(p_i) with

$$\log(p_i/(1-p_i)) = (\alpha_i + \sum_{j\in\partial i} \theta_{ij} x_j)$$

- a very useful model for dependent binary variables (*NB various parameterisations*)

Statistical mechanics models

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

The classic Ising model (for ferromagnetism) is the symmetric autologistic model on a square lattice in 2-D or 3-D. The Potts model is the generalisation to more than 2 ‘colours’

$$p(X) \propto \exp\left\{\sum_i \alpha_{x_i} + \beta \sum_{i\sim j} I[x_i = x_j]\right\}$$

and of course you can usefully un-symmetrise this.

Auto-Poisson model

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

$$\propto \exp\left\{\sum_i (\alpha_i x_i - \log x_i!) + \sum_{i\sim j} \theta_{ij} x_i x_j\right\}$$

$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$

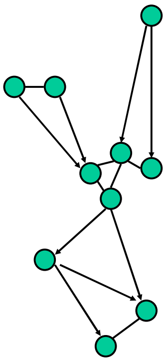
is Poisson $(\exp(\alpha_i + \sum_{j\in\partial i} \theta_{ij} x_j))$

For integrability, θ_{ij} must be ≤ 0, so this only models negative dependence: very limited use.

Hierarchical models and hidden Markov processes

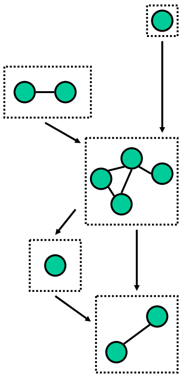
Chain graphs

- If both directed and undirected edges, but no directed loops:
- can rearrange to form global DAG with undirected edges within blocks



Chain graphs

- If both directed and undirected edges, but no directed loops:
- can rearrange to form global DAG with undirected edges within blocks
- Hammersley-Clifford within blocks

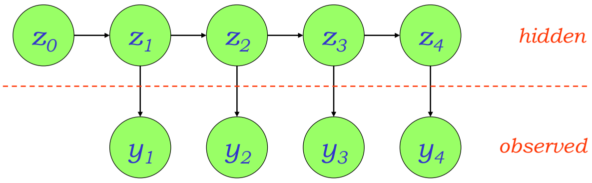


Hidden Markov random fields

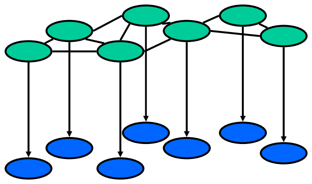
- We have a lot of freedom modelling spatially-dependent continuously-distributed random fields on regular or irregular graphs
- But very little freedom with discretely distributed variables
- ⇒ use hidden random fields, continuous or discrete
- compatible with introducing covariates, etc.

Hidden Markov models

e.g. Hidden Markov chain



Hidden Markov random fields



Unobserved dependent field

Observed conditionally-independent discrete field

(a chain graph)

Spatial epidemiology applications

relative risk

expected cases

cases $Y_i \sim Poisson(\lambda_i e_i)$

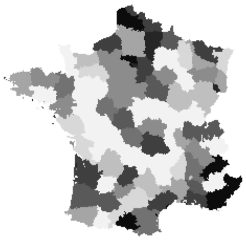
independently, for each region *i*. Options*:*

- $\log(\lambda_i)$ CAR, CAR+white noise (BYM, 1989)
- Direct modelling of $\operatorname{cov}(\log(\lambda_i))$,e.g. SAR
- Mixture/allocation/partition models:

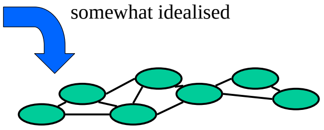
$$E(Y_i) = \lambda_{z_i} e_i$$

- Covariates, e.g.: $E(Y_i) = \lambda_{z_i} e_i \exp(x_i^T \beta)$

Spatial epidemiology applications



Spatial contiguity is usually somewhat idealised



Spatial epidemiology applications

Richardson & Green (*JASA*, 2002) used a hidden Markov random field model for disease mapping

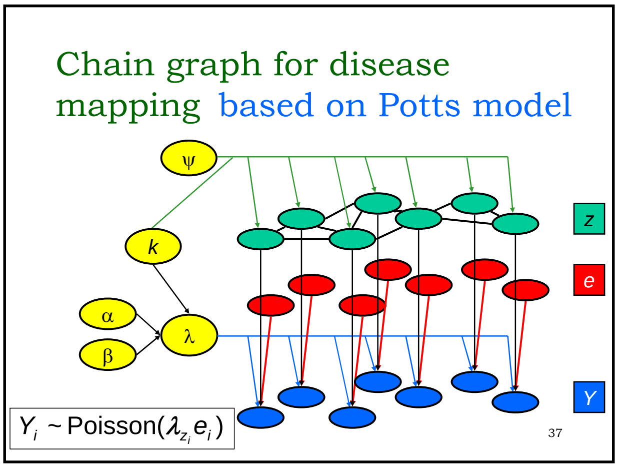
$$y_i \sim \text{Poisson}(\lambda_{z_i} e_i)$$

observed incidence

relative risk parameters

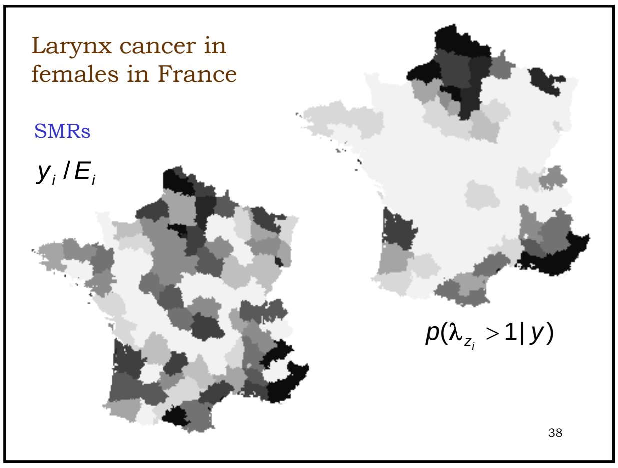
hidden MRF

expected incidence



Chain graph for disease mapping based on Potts model

$Y_i \sim \text{Poisson}(\lambda_{z_i} e_i)$



Larynx cancer in females in France

SMRs

y_i / E_i

$p(\lambda_{z_i} > 1 | y)$